AMERICAN HEALTH LAW ASSOCIATION

HEALTH LAW WEEKLY

- April 03, 2026

Who Owns the Pulse? What Health System Lawyers Need to Know About Licensing Patient Data in the Age of AI

- **Mehrin ("Mir") Masud-Elias**, Arnold & Porter

A landmark paper published in *Science* in March 2026[1] posed a question health system lawyers can no longer treat as academic: should real-world health data be governed as a public utility—essential infrastructure, like electricity or clean water, subject to enforceable standards, transparent economics, and meaningful patient input? As the authors of the paper observe, health data infrastructure today is largely voluntary, fragmented, and economically misaligned—and the consequences of that misalignment are becoming impossible to ignore.

Whether or not the public utility model ultimately takes hold, the governance challenge it describes is already here. Clinical records, imaging data, genomic datasets, and real-world evidence now power AI systems supporting drug discovery, clinical decision support (including trial enrollment), workflow optimization, and regulatory submissions. Hospitals and health systems sit at the center of this transformation—as data sources, AI deployers, and licensing counterparties in complex commercial arrangements their legacy legal frameworks were never designed to govern.

For health system counsel, patient data licensing is no longer a narrow privacy compliance question. It is a multi-disciplinary governance challenge implicating the Health Insurance Portability and Accountability Act (HIPAA), a rapidly expanding stack of state AI laws, Food and Drug Administration (FDA) regulatory strategy, intellectual property, cybersecurity, and most importantly, community trust. This article offers a practical map to navigate these choppy waters.

Patient Data Licensing is a Risk Transfer Transaction

Every patient data licensing arrangement is, at its core, a structured transfer of legal, reputational, and ethical risk. Institutions treating these arrangements as routine administrative agreements—governed by legacy DUA templates and baseline HIPAA analysis—are systematically underestimating their legal exposure.

The threshold question is not "Can we share this under HIPAA?" It is "Who governs this data, and for whose benefit?" That reframing has significant practical consequences. "Patient data licensing" encompasses a wide range of arrangements with distinct risk profiles: AI model training raises derivative ownership questions simple research data sharing does not; a validation dataset may implicate FDA regulatory strategy in ways a de-identified analytics license will not; remuneration arrangements require "sale-of-PHI" analysis; federated learning reduces raw data exposure but does not eliminate governance obligations. Classification drives risk—the same data may carry dramatically different compliance obligations depending entirely on how the arrangement is structured. Same with whether or not the patient-derived data is fully or partially de-identified under HIPAA, or otherwise.

In this brave new world, compliance officers, privacy counsel, and research governance professionals must be at the table from the outset, not brought in after agreements have been substantially negotiated.

Ownership, Control, and Governance: Getting the Framing Right

One of the most persistent misconceptions in health care AI transactions is the conflation of ownership, control, and governance.

Institutions often speak loosely about "owning" patient data. This framing is legally inaccurate. In most U.S. jurisdictions, the physical and electronic medical record is owned by the institution that created it—not the patient. Patients hold important *rights with respect to* their records (access, amendment, and accounting of disclosures under HIPAA), but those are regulatory entitlements, not property rights in the traditional sense. Institutions are stewards, not proprietors: their authority over patient data is bounded by HIPAA, state law, IRB requirements, and consent terms that collectively constrain how that data can be used. Health systems that negotiate AI licensing arrangements as if they have unlimited authority over patient data as its "owners" are miscalculating their legal position—and their governance obligations.

Institutional *control*—exercised through BAAs, DUAs, and access management systems—is real but limited. Once a model is trained on institutional data, the institution's ability to govern downstream use is limited to what the contract specifies. Fine-tuned model weights, embeddings, and intermediate

artifacts may be entirely beyond the reach of post-hoc control mechanisms. This is why contractual drafting is so consequential—and why governance must precede, not follow, the transaction.

Governance is the emerging priority. Governance asks: who decides, on what standards, with what accountability structures, and for whose benefit? Every practical recommendation that follows flows from that organizing question.

Data Classification Is Not a One-Time Checkbox

The regulatory analysis of any licensing arrangement begins with a threshold question: what type of data is involved? AI systems typically do not consume a single data type in isolation—multi-modal architectures routinely combine electronic health records (EHRs) with imaging, genomics, claims, and social determinants of health data, and the combined dataset may trigger legal obligations that no individual component would generate alone. This combinatorial risk is one of the central challenges in AI data compliance.

Several categories deserve particular attention. Medical imaging frequently contains embedded PHI in DICOM metadata not removed by standard de-identification workflows—institutions often discover this only after a request is in progress. Consumer-generated health data from wearables falls into a significant and growing HIPAA gap, governed instead by state consumer privacy laws like Washington's My Health My Data Act. Genomic data is subject to state genetic privacy statutes that vary significantly and can be triggered even when data appears de-identified. Social and external data—housing, income, environmental factors—creates heightened re-identification risk when combined with clinical data at scale.

Data classification must be assessed across four dimensions—identifiability, consent, applicable law, and re-identification exposure—and revisited as AI systems evolve and datasets are combined. It is a dynamic obligation, not a one-time transaction checkpoint.

The Regulatory Stack: HIPAA is the Floor, Not the Ceiling

A single AI licensing arrangement may simultaneously implicate HIPAA, multiple state privacy and genetic statutes, FDA device regulation, EU AI Act obligations, and FTC algorithmic bias enforcement authority. The compliance task is to understand how these frameworks interact—and where they conflict.

HIPAA's Three Pathways

- Safe Harbor—removal of 18 specified identifiers—is the default choice because it requires no statistical expertise, but it is poorly suited for AI applications. Dates more specific than year and geographic data below state level must be removed, foreclosing the longitudinal and geographic granularity that model training typically requires.
- Expert Determination—statistical analysis establishing that re-identification risk is very small—is more complex and expensive, but it is the appropriate standard for AI licensing transactions. Dates and geographic granularity can be retained where risk is statistically acceptable, and a well-documented Expert Determination opinion is one of the most valuable assets in a

regulatory inquiry or litigation context. The FDA's emerging lifecycle-based oversight framework and EU conformity assessments increasingly expect Expert Determination-level documentation.

- Limited Data Sets (LDS) remain PHI, with all HIPAA obligations attached. The critical gap: standard DUA language—including HHS model language—does not address model training, derivative ownership, fine-tuned weight restrictions, or model lineage. Every AI-related LDS transaction requires bespoke drafting, and fee arrangements require analysis under HIPAA's "sale-of-PHI" prohibition regardless of how they are characterized.

State law

Texas TRAIGA (effective January 1, 2026) focuses primarily on prohibiting specific harmful AI uses—behavioral manipulation, discrimination, and exploitation of minors—through an intent-based liability framework, and requires licensed health care practitioners to disclose AI use in diagnosis or treatment; notably, the enacted version is significantly narrower than earlier drafts and does not impose a general affirmative patient consent requirement. California's AB 3030 and SB 1120 (both effective January 1, 2025) impose disclosure requirements for AI-generated patient communications and human oversight mandates for AI in utilization review, respectively; AB 2013 (compliance deadline January 1, 2026) requires AI developers to publicly disclose training data sources, creating direct tension with trade secret strategy. Illinois prohibits autonomous AI therapy. Washington's My Health My Data Act covers consumer-generated health data with a private right of action. Colorado's AI Act (effective June 2026) requires impact assessments for high-risk AI systems. A December 2025 executive order signals potential federal preemption, but state laws remain enforceable today.

International obligations

For health systems with any EU exposure, the EU AI Act is no longer a horizon risk. The full high-risk AI compliance framework—which expressly covers health care AI analyzing patient-specific information—takes effect August 2, 2026, with extraterritorial reach whenever AI outputs are used in the EU. Many U.S. institutions are behind on conformity assessments. Under GDPR, health data requires explicit consent or a specific derogation; standard U.S. research consent language frequently falls short. The European Health Data Space, in force since 2025, creates a secondary use framework that will increasingly interact with AI licensing arrangements.

AI as a Regulated Medical Device: Risk Is Not a Deployment Event

AI systems that analyze patient-specific information and generate clinical recommendations may constitute Software as a Medical Device (SaMD) under the FDCA. Intended use drives classification—advisory tools may fall outside SaMD scope, while systems producing autonomous clinical outputs are far more likely to require FDA clearance. Intended use must be documented from day one, not characterized retroactively. Risk spans the entire model lifecycle, from data collection through post-market performance. At data collection: consent, identifiability, and bias in source populations. Licensing agreements must be designed for this full lifecycle.

The Federated Learning Governance Gap

Federated learning—in which raw data never leaves the institution and only model parameters are shared—reduces re-identification risk and may simplify IRB and consent analysis in some contexts. But "data stays with us" does not mean governance questions go away. Model inversion attacks can reconstruct training data from shared parameters; membership inference attacks can identify individuals in the training set. Parameter ownership and downstream model rights still require explicit contractual treatment regardless of architectural choice. Federated architectures change the risk profile—they do not eliminate it.

The Governance Question: "Should We?" is Separate from "Can We?"

Regulatory compliance answers what is legally permissible. Governance answers what is institutionally appropriate. An arrangement may be technically compliant and still be inconsistent with institutional mission, community expectations, or long-term reputational interests, especially for tax-exempt entities.

The old framing—"Is this de-identified?" "Does our DUA permit this?"—is no longer adequate. The new framing regulators and communities are applying is: "How is patient data generating economic value—and have we governed that responsibly?" and "Would our patients recognize their interests in this arrangement?" Academic medical centers operating under tax-exempt status must affirmatively answer: how does this arrangement serve our community?

Existing consent authorizations were typically drafted before commercial AI training was a common downstream use. Research governance professionals should audit existing consent language and assess whether it covers the proposed AI use—and if not, whether new consent is ethically required even if not legally mandated.

Community trust is a material governance risk, not a soft consideration. Research consistently shows patients are willing to share data for purposes they understand and endorse—and significantly less willing when commercial benefit flows primarily to third parties. And once data is licensed, the institution's reputation travels with it through every downstream use.

What Health System Counsel Should Be Doing Now

Get the contract right—the first time

AI data licensing agreements require contractual architecture specifically designed for the AI use case—not adapted from legacy data sharing or research collaboration templates. Each clause should reflect a deliberate allocation of risk, and the contract must be evaluated as an integrated whole—a narrowly drawn field of use clause is ineffective if the scope of use clause is broad, and a robust termination provision is undermined if it does not address the fate of fine-tuned weights.

The definition of data clause is often the most consequential and most underspecified provision. Curated, annotated, or enriched datasets embody substantial institutional intellectual contribution that should be reflected in the economic terms—not treated as equivalent to an unprocessed data transfer.

Ownership must be addressed across the full AI stack: inputs, outputs, intermediate artifacts, and fine-tuned weights. The fine-tuned weights are the factory—they encode the productive output of training on institutional data and carry substantial commercial value. That question must be answered in the

contract, not litigated after the fact. Four AI-specific provisions are most frequently absent and most likely to generate disputes: training restrictions, model lineage recordkeeping, audit rights through the full model lifecycle, and survival provisions upon termination.

Treat bias testing as an enforcement obligation, not a best practice

Contracts should require demographic stratification of validation results, explainability documentation sufficient for regulatory review, context-specific validation metrics, lifecycle drift monitoring with defined revalidation triggers, and audit rights extending to bias testing results. California law and the EU AI Act both require algorithm inspection and post-market monitoring—these are enforceable today.

Capture the full economic value

Institutions frequently undervalue their data. Indirect monetization channels—AI validation arrangements, workflow optimization data, embedded analytics—generate ongoing commercial value that flat fees fail to capture. Counsel should advocate for royalty structures tied to commercial success. Revenue allocation policies aligned with institutional mission are also a community trust imperative: the *Science* paper frames directly that patients currently have no formal role in governing how their data generates economic value.

IP Strategy

AI-enabled development requires a hybrid intellectual property strategy that must be designed from the outset, not assembled after disputes arise. Training datasets, model weights, feature engineering methodologies, and negative data—what didn't work—are frequently protected as trade secrets, requiring robust confidentiality controls throughout the model lifecycle. Patents may protect training methods and processes, methods of AI-enabled clinical use, and device integration methods. Institutions should clearly define at the contracting stage what they are contributing, what they are retaining, and what the counterparty may and may not do with derivatives—because once a model is trained, the IP boundaries that were not drawn in the contract will be litigated instead.

Governance and Compliance Checklists

Effective governance is an operational system, not a policy document. The institutions navigating this landscape most effectively have invested in cross-functional governance infrastructure built on five components: cross-functional committees (privacy, security, legal, research compliance, and clinical leadership); tiered risk scoring calibrated to data sensitivity and AI intended use; defined escalation pathways for executive or board-level review; audit-ready documentation of governance decisions and de-identification methodologies; and local vendor oversight requiring validation in the institution's own patient population context.

The governance and compliance obligations described above translate into a concrete set of institutional requirements. Health system counsel should assess their institution's current state against each of the following—and treat any gap as a priority:

- Data inventory and classification—AI-specific use categories, combinatorial re-identification risk, intended-use documentation

- Regulatory pathway alignment—State law mapping; EU AI Act high-risk classification review
- Governance committee review—Cross-functional structure, tiered risk scoring, escalation pathways, audit-ready documentation
- AI-specific contract drafting—Full AI stack ownership; training restrictions; model lineage; survival provisions; AI-specific cybersecurity
- Bias validation obligations—Demographic stratification; explainability; context-specific metrics; lifecycle drift monitoring
- Consent language audit—Adequacy for AI training, derivative ownership, and commercialization; IRB assessment
- Economic value analysis—Direct and indirect monetization; royalty structures; revenue allocation policies
- IP strategy review—Trade secret and patent analysis

Conclusion

The old compliance framing—"Is this de-identified?" "Does our DUA permit this?"—is no longer adequate for the decisions health system lawyers are being asked to support. The question regulators, communities, and institutional leaders are increasingly asking is more demanding: how is patient data generating economic value, and have we governed that responsibly?

As Haendel et al. argue in *Science*, to fully realize precision medicine, drug safety, and population health for everyone, real-world health data must be treated as an essential resource—not a private asset extracted without accountability to the patients and communities who generated it. Health system lawyers are uniquely positioned to help their institutions close that governance gap. The institutions best positioned to manage risk in the AI era will be those that recognized early that patient data licensing is a risk transfer transaction—and built governance infrastructure equal to its complexity.

About the Author

Mehrin ("Mir") Masud-Elias is Counsel at Arnold & Porter in Washington, D.C. and Boston, where she advises clients across life sciences, healthcare, tax-exempt organizations, and academic research. She previously served as Head of Legal at the University of Pennsylvania's Abramson Cancer Center. The views expressed in this article are solely those of the author and do not constitute legal advice.

[1] Melissa A. Haendel et al., "Governing Real-World Health Data as a Public Utility," *Science* 391, 993–996 (2026). DOI: 10.1126/science.aeb1178.